(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0049500**

(51)[2022.01] **H04L 43/062**; H04L 47/24

(13) **B**

(21) 1-2023-06051

(22) 08/09/2023

(45) 25/08/2025 449

(43) 25/12/2023 429A

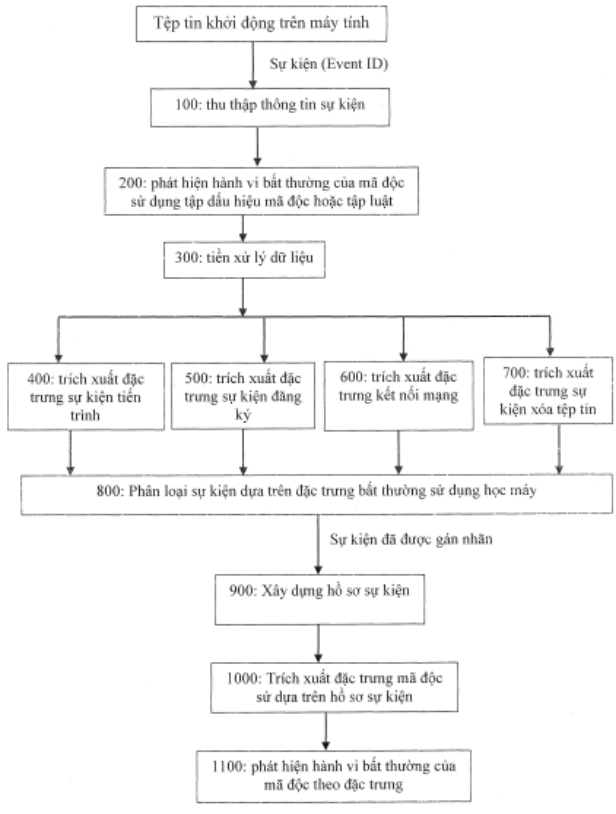
(76) Đỗ Xuân Chợ (VN)
Số nhà 12A03, sảnh B, chung cư FLC Landmark Tower, Lê Đức Thọ, Mỹ Đình 2, Nam Từ Liêm, Hà Nội

(74) CÔNG TY CỔ PHẦN TƯ VẤN BIGPRO (BIGPRO CONSULTATION JOIN STOCK)

(54) PHƯƠNG PHÁP PHÁT HIỆN MÃ ĐỘC

(21) 1-2023-06051

(57) Sáng chế đề cập đến phương pháp phát hiện mã độc bao gồm: thu thập, lưu trữ, bởi công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng, các dấu hiệu có trong các sự kiện gắn với tệp tin; so khớp các dấu hiệu đã nêu với tập dấu hiệu mã độc hoặc kiểm tra các dấu hiệu trong các sự kiện gắn với tệp tin đã nêu bằng tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc đã biết trước đó. Nếu kết quả là có mã độc thì có thể phát hiện được rất nhanh chóng và đơn giản. Trái lại, nếu kết quả là không có mã độc thì tiếp tục thực hiện việc phát hiện mã độc bằng cách trích xuất các đặc trưng theo các sự kiện gồm có ít nhất là sự kiện tiến trình, sự kiện đăng ký, sự kiện kết nối mạng, và sự kiện xóa tệp tin. Các đặc trưng theo sự kiện này được tổng hợp thành dạng đồ thị biểu diễn mối quan hệ sự kiện theo thời gian giữa chúng. Đồ thị này được trích xuất để thu được các các đặc trưng về sự kiện theo thời gian gắn với tệp tin. Cuối cùng, mô hình học máy sẽ được sử dụng để phát hiện mã độc dựa trên đầu vào là các đặc trưng về sự kiện theo thời gian đã nêu.



Hình 1

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp phát hiện mã độc, cụ thể hơn là đến phương pháp phát hiện mã độc thông qua các hành vi bất thường của mã độc dựa trên các chuỗi sự kiện theo tiến trình hoặc theo thời gian được ghi nhận, phát hiện hoặc dự đoán có hay không mã độc liên quan đến tệp tin tương ứng thông qua các đặc trưng được trích xuất trong chuỗi sự kiện.

Tình trạng kỹ thuật của sáng chế

Như đã biết, tấn công vào hệ thống bằng cách phát tán mã độc là kỹ thuật rất nguy hiểm và ngày càng được phổ biến vì nó tấn công trực tiếp vào điểm yếu của con người trong hệ thống. Chính vì lý do này, phát hiện mã độc đang trở thành một nhiệm vụ cấp thiết cho các hệ thống đảm bảo an toàn thông tin. Hiện nay đang có hai hướng tiếp cận để phát hiện hành vi bất thường của mã độc. Hướng tiếp cận đơn giản nhất là phát hiện hành vi bất thường của mã độc dựa vào lưu trữ và phân tích các mã độc đã bị phát hiện. Các đặc điểm của mã độc như mã băm của mã độc (mã hash), tên miền, IP (Internet Protocol), v.v., sẽ được so sánh, đối chiếu với các đặc điểm về mã độc đã có trong cơ sở dữ liệu được thu thập thông qua quá trình lưu trữ và phân tích các loại mã độc đã được phát hiện ra trước đó. Theo cách này, có thể phát hiện nhanh chóng và chính xác các hành vi bất thường của mã độc đối với các loại mã độc lặp lại hoặc có tính chất tương tự với các mã độc đã được lưu trữ và phân tích. Tuy nhiên, cách này được cho là khó để phát hiện được các hành vi bất thường hoặc các dấu hiệu chưa được biết tới. Bởi vì trong quá trình thiết kế và phát triển mã độc, những kẻ tấn công thường sử dụng các phương pháp mới, các hành vi mới. Vì thế, sẽ gây ra rất nhiều khó khăn cho các công cụ giám sát và phát hiện mã độc.

Để giải quyết các nhược điểm của phương pháp phát hiện theo dấu hiệu, các hướng tiếp cận gần đây đã không tìm cách so sánh và đối chứng từng dấu hiệu

đơn lẻ nữa. Thay vào đó, các hành hành vi bất thường của mã độc sẽ được liên kết với nhau thành các chuỗi sự kiện. Chuỗi sự kiện này thể hiện mối quan hệ và một số quy tắc tấn công và phát tán của các loại mã độc đã biết. Sau đó, chuỗi sự kiện thể hiện hành vi bất thường của mã độc sẽ được so sánh với các tập luật đã được xây dựng từ các loại mã độc đã biết. Tập luật này chính là các chiến lược tấn công cũng như các phương pháp khai thác hệ thống của mã độc. Như vậy có thể thấy rằng, với hướng tiếp cận này thì mặc dù các dấu hiệu thu được của mã độc có thể khác biệt hoàn toàn với các dấu hiệu của mã độc đã có trong hệ thống. Tuy nhiên, chỉ cần các sự kiện được ghi nhận trùng hoặc giống với các sự kiện đã được ghi nhận từ mã độc trong tập luật thì hệ thống vẫn có thể dễ dàng đưa ra kết luận là, ví dụ có mã độc hoặc có mã độc liên quan đến tệp tin tương ứng.

Các hướng tiếp cận phát hiện mã độc sử dụng học máy đã được đề xuất trong thời gian gần đây thường thực hiện theo cách thức là sử dụng Sanbox (một công cụ ảo hóa) để thực hiện thu thập hành vi bất thường của mã độc. Sau đó sẽ tiến hành trích xuất các hành vi bất thường của mã độc dựa trên các vết nhật ký của Sanbox. Tuy nhiên, theo cách tiếp cận này vì loại mã độc thường có chức năng nhận diện và chống lại Sanbox, ngủ đông, v.v.. Chính vì vậy các thuộc tính mã độc được thu thập từ nhật ký Sanbox thường không mang nhiều ý nghĩa. Bên cạnh đó, trong quá trình tấn công, mã độc thường sử dụng nhiều kỹ thuật khai thác và đánh cắp dữ liệu tại những mốc thời gian khác nhau nhằm che giấu các hành vi bất thường của chúng. Do đó các phương pháp trích xuất thuộc tính truyền thống sẽ không thể tính toán và trích xuất được các thuộc tính thể hiện sự liên hệ giữa các hành vi của mã độc. Điều này dẫn đến các hệ thống phát hiện mã độc sẽ cho rằng các hành vi được ghi nhận đó là lành tính mặc dù chúng đều là những hành vi che giấu và ẩn mình của mã độc. Chính vì vậy, với hướng tiếp cận truyền thống thì không chỉ thiếu hành vi của mã độc mà các hành vi này còn không thể hiện được sự tương quan và mối quan hệ của chúng với nhau.

Do đó, có nhu cầu về giải pháp phát hiện mã độc hiệu quả hơn, có khả năng phát hiện tốt các loại mã độc khác nhau, có tính đến nhiều đặc trưng đa dạng một cách tổng thể, để hỗ trợ hiệu quả cho việc đảm bảo an toàn của hệ thống máy tính.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp phát hiện mã độc hoặc hành vi bất thường liên quan đến mã độc, khắc phục được một hoặc một số vấn đề còn tồn tại nêu trên.

Một mục đích khác của sáng chế là đề xuất phương pháp phát hiện mã độc, có tính đến nhiều đặc trưng đa dạng khác nhau một cách tổng thể, để hỗ trợ hiệu quả cho việc giám sát hành vi bất thường liên quan đến mã độc.

Ngoài ra, sáng chế cũng đề xuất phương pháp phát hiện hành vi bất thường liên quan đến mã độc kế thừa được một hoặc một số các ưu điểm của các phương pháp phát hiện mã độc đã biết.

Một mục đích khác nữa của sáng chế là đề xuất phương tiện lưu trữ có chứa chương trình máy tính để thực thi phương pháp phát hiện mã độc.

Cần hiểu rằng, sáng chế không bị giới hạn ở các mục đích được nêu ra trên đây, mà còn bao gồm các mục đích khác nữa mà có thể được nhận biết bởi người có hiểu biết trung bình về lĩnh vực kỹ thuật tương ứng trong các mô tả dưới đây.

Để đạt được một hoặc một số các vấn đề nêu trên, sáng chế đề xuất phương pháp phát hiện mã độc bao gồm:

thu thập, bởi công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng, các dấu hiệu có trong các sự kiện gắn với tệp tin, các sự kiện gắn với tệp tin này bao gồm ít nhất là sự kiện tiến trình (Event process), sự kiện đăng ký (Registry Event), sự kiện kết nối mạng (Event Network connection), sự kiện xóa tệp tin (Event File Delete);

so khớp các dấu hiệu trong các sự kiện đã nêu với tập dấu hiệu mã độc hoặc kiểm tra các dấu hiệu trong các sự kiện gắn với tệp tin đã nêu bằng tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc, để thực hiện việc phát hiện mã độc theo cấp độ thứ nhất xem liệu có mã độc liên quan tới tệp tin tương ứng hay không,

trong đó tập dấu hiệu mã độc và tập luật phát hiện mã độc dựa trên các dấu

hiệu về hành vi bất thường của mã độc đã có trong cơ sở dữ liệu, được tổng hợp từ các mã độc đã biết trước đó, và tập luật phát hiện mã độc được xây dựng nhờ sử dụng MITRE ATT&CK (Adversarial Tactics Techniques & Common Knowledge) Framework;

nếu kết quả là không có mã độc thì tiếp tục thực hiện việc phát hiện mã độc theo cấp độ thứ hai, trong đó việc phát hiện mã độc theo cấp độ thứ hai được thực hiện thông qua các bước sau:

trích xuất các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình (Event process), các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký (Registry Event), các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng (Event Network connection), và các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin (Event File Delete),

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về tiến trình đã nêu để phát hiện các tiến trình bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên (Random Forest) đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về đăng ký đã nêu để phát hiện các đăng ký bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về kết nối mạng đã nêu để phát hiện các kết nối mạng bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về xoá tệp tin đã nêu để phát hiện các tiến trình xoá tệp tin bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

tổng hợp ít nhất là các tiến trình bất thường, các đăng ký bất thường, các kết nối mạng bất thường, và các tiến trình xoá tệp tin bất thường đã nêu thành dạng đồ thị biểu diễn mối quan hệ sự kiện theo thời gian giữa chúng,

trích xuất, sử dụng mạng đồ thị học sâu, các đặc trưng về sự kiện theo thời gian gắn với tệp tin dựa trên đồ thị biểu diễn mối quan hệ sự kiện theo thời gian đã nêu,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về sự kiện theo thời gian đã nêu để phát hiện mã độc theo cấp độ thứ hai xem liệu có mã độc liên quan đến tệp tin tương ứng hay không, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Theo một phương án, các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình, bao gồm nhưng không giới hạn ở, loại tiến trình; loại tệp tin; số lượng tệp mà tiến trình tạo hoặc ghi đè; và tiến trình có mức ưu tiên truy cập thấp hay không.

Theo một phương án, các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký, bao gồm nhưng không giới hạn ở, số lượng đăng ký sử dụng; số lượng HKEY_LOCAL_MACHINE; số lượng HKEY_CLASSES_ROOT; có sử dụng các khoá đăng ký khởi tạo hay không; và có sử dụng các khoá đăng ký cài đặt đang hoạt động hay không.

Theo một phương án, các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng, bao gồm nhưng không giới hạn ở, tiến trình có kết nối mạng hay không; số lần kết nối mạng của tiến trình; địa chỉ IP, tên miền hoặc yêu cầu HTTP mà tiến trình kết nối tới; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không rõ nguồn gốc; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm có dấu hiệu nghi ngờ; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không hợp pháp; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm đáng tin cậy; và số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không an toàn.

Theo một phương án, các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin, bao gồm nhưng không giới hạn ở, số lượng các tệp tin tiến trình xóa; tổng dung lượng các tệp tin tiến trình xóa; có tiến trình xóa tệp tin ở ổ người dùng hay không; có tiến trình xóa tệp tin ở ổ $RECYCLE.BIN

hay không; có tiến trình xóa tệp tin ở ổ chứa hệ điều hành hay không; có tiến trình xóa tệp tin ở ổ chứa dữ liệu hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình (x86) hay không; có tiến trình xóa tệp tin ở ổ chứa các Perflog hay không; có tiến trình xóa tệp tin ở ổ chứa thư mục OneDriveTemp hay không; có tiến trình xóa tệp tin ở ổ chứa $WinREAgent hay không; số lượng tệp tin xml tiến trình xóa; số lượng tệp tin ảnh tiến trình xóa; số lượng tệp tin HTML tiến trình xóa; số lượng tệp tin văn bản tiến trình xóa; và số lượng tệp tin thực thi tiến trình xóa.

Phương pháp nêu trên có thể còn bao gồm phân loại các sự kiện dựa trên các đặc trưng được xác định trước có trong tệp thực thi, và tập hợp các đặc trưng được xác định trước có trong tệp thực thi này vào đồ thị biểu diễn mối quan hệ sự kiện theo thời đã nêu.

Công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng là một trong số công cụ thu thập và lưu trữ sự kiện Sysmon, Wahzu, hoặc Endpoint Security được cài đặt theo gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu.

Theo một khía cạnh khác, sáng chế đề xuất phương tiện lưu trữ có chứa chương trình máy tính bao gồm các lệnh mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp nêu trên.

Mô tả vắn tắt các hình vẽ

Hình 1 là sơ đồ khối thể hiện phương pháp phát hiện hành vi bất thường của mã độc theo một phương án ưu tiên của sáng chế;

Hình 2 là giản đồ giản lược thể hiện việc huấn luyện và sử dụng mô hình học máy trong việc phát hiện các sự kiện bất thường trong tệp tin theo sáng chế;

Hình 3 là giản đồ giản lược thể hiện ví dụ về hồ sơ hành vi mã độc thể hiện mối quan hệ giữa các sự kiện dạng cây theo một phương án ưu tiên của sáng chế; và

Hình 4 là giản đồ giản lược thể hiện việc huấn luyện và sử dụng mô hình

học máy trong việc phát hiện các hành vi bất thường trong tệp tin theo sáng chế.

Mô tả chi tiết sáng chế

Dưới đây, các ưu điểm, hiệu quả và bản chất của sáng chế có thể được hiểu rõ hơn thông qua việc mô tả chi tiết các phương án ưu tiên thực hiện có dựa vào các hình vẽ kèm theo. Trên các hình vẽ, các số chỉ dẫn giống nhau được dự định để biểu thị các thành phần hoặc chi tiết giống nhau hoặc tương đương và được sử dụng thống nhất trong toàn bộ mô tả, do vậy trên một số các hình vẽ hoặc một số phần trên hình vẽ có thể không xuất hiện một hoặc một số các số chỉ dẫn nhằm mục đích làm cho hình vẽ trở nên đơn giản và thuận tiện trong việc thể hiện các thành phần cấu tạo hoặc nguyên lý hoạt động khác nhau của sáng chế, trong trường hợp như vậy, mối tương quan giữa các thành phần hoặc chi tiết cụ thể với các số chỉ dẫn biểu thị nó có thể được minh họa rõ khi tham chiếu tới các hình vẽ khác hoặc các phần khác trên hình vẽ. Bên cạnh đó, các hình vẽ có thể được thể hiện theo các cách khác nhau với dự định để phác họa khái niệm hoặc nguyên lý theo sáng chế, và do đó các thành phần và chi tiết được thể hiện trên hình vẽ là không theo kích thước và hình dạng thực tế, một số thành phần hoặc chi tiết sẽ được phóng đại lên và có thể được biểu thị bởi các khối giản lược nhằm mục đích minh họa và thuận tiện cho việc mô tả. Vì vậy, cần hiểu rằng các phương án được mô tả trong phần mô tả chỉ với mục đích làm ví dụ giúp cho việc hiểu rõ hơn về bản chất và các ưu điểm của sáng chế, mà không giới hạn phạm vi của sáng chế theo các phương án được mô tả này.

Trên Hình 1 là sơ đồ khối thể hiện phương pháp phát hiện hành vi bất thường của mã độc theo một phương án ưu tiên của sáng chế.

Như được thể hiện trên Hình 1, phương pháp phát hiện hành vi bất thường của mã độc theo phương án ưu tiên này bao gồm các bước từ bước 100 đến bước 1100 như được mô tả chi tiết hơn dưới đây.

Bước 100: thu thập thông tin sự kiện.

thu thập, bởi công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng, các dấu hiệu có trong các sự kiện gắn với tệp tin, các sự kiện gắn

với tệp tin này bao gồm ít nhất là sự kiện tiến trình (Event process), sự kiện đăng ký (Registry Event), sự kiện kết nối mạng (Event Network connection), sự kiện xóa tệp tin (Event File Delete).

Tuỳ ý, công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng là một trong số công cụ thu thập và lưu trữ sự kiện Sysmon, Wahzu, Endpoint Security, hoặc các công cụ tương tự được cài đặt theo gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu.

Bước 200: phát hiện hành vi bất thường của mã độc sử dụng tập dấu hiệu mã độc hoặc tập luật.

Ở bước này sẽ so khớp các dấu hiệu trong các sự kiện đã nêu với tập dấu hiệu mã độc hoặc kiểm tra các dấu hiệu trong các sự kiện gắn với tệp tin đã nêu bằng tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc, để thực hiện việc phát hiện mã độc theo cấp độ thứ nhất xem liệu có mã độc liên quan tới tệp tin tương ứng hay không. Tập dấu hiệu mã độc và tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc đã có trong cơ sở dữ liệu, được tổng hợp từ các mã độc đã biết trước đó, và tập luật phát hiện mã độc được xây dựng nhờ sử dụng MITRE ATT&CK (Adversarial Tactics Techniques & Common Knowledge) Framework.

Các thông tin về dấu hiệu và hành vi bất thường của mã độc được thu thập tại các máy tính người dùng sử dụng dưới dạng các sự kiện. Các sự kiện này được phân tích để xem liệu có bất thường nào giống với hành vi bất thường của mã độc hay không dựa trên tập các dấu hiệu về hành vi bất thường của mã độc có trong sự kiện hoặc tập luật. Khi phát hiện có một, một số, hoặc một nhóm các dấu hiệu mã độc (có thể gọi chung là tập luật) trong sự kiện đang giám sát là có sự trùng khớp hoặc tương tự với một, một số, hoặc một nhóm dấu hiệu trong số các dấu hiệu về hành vi bất thường của mã độc được tổng hợp từ các mã độc đã biết trước đó thì có thể nhận biết ngay được sự bất thường trong tệp tin này hay nói cách khác là phát hiện được mã độc liên quan đến tệp tin này. Ở đây, tập luật bất thường dựa trên các dấu hiệu bất thường đã có trong cơ sở dữ liệu, được tổng hợp từ các

mã độc đã biết trước đó, từ các nền tảng mã nguồn mã, từ tri thức cộng đồng, hoặc từ những nguồn thích hợp bất kỳ.

Theo một phương án ưu tiên, việc thu thập các sự kiện trên máy người dùng có thể sử dụng công cụ thu thập dữ liệu Wahzu, công cụ Wahzu được cho là có thể thu thập khá đầy đủ toàn bộ các sự kiện do tệp tin sinh ra, ví dụ sự kiện tiến trình (Event process), sự kiện đăng ký (Registry Event), sự kiện kết nối mạng (Event Network connection), sự kiện xóa tệp tin (Event File Delete).

Một ví dụ về công cụ Wahzu là công cụ thu thập các sự kiện trên máy người dùng được cài đặt theo gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu, Windows, v.v.. Công cụ mạng Wahzu có thể cho phép tự thu thập các sự kiện được sinh ra từ hệ điều hành máy tính và các ứng dụng khác, từ đó có thể thu thập được tiến trình, sự kiện của tệp tin như sự kiện tiến trình (Event process), sự kiện đăng ký (Registry Event), sự kiện kết nối mạng (Event Network connection), sự kiện xóa tệp tin (Event File Delete), v.v., phục vụ cho việc giám sát và phát hiện hành vi bất thường của mã độc trong tệp tin.

MITRE ATT&CK Framework có thể được coi là giải pháp tổng thể phát hiện mã độc dựa trên các sự kiện và hành vi bất thường của chúng. MITRE ATT&CK Framework bao gồm hai thành phần chính là chiến lược và chiến thuật của mã độc. Từ chiến lược và chiến thuật này có thể xây dựng được các tập luật phục vụ cho việc giám sát và phát hiện hành vi bất thường của sự kiện trong tệp tin chứa mã độc.

Theo cách này, có thể phát hiện nhanh chóng và chính xác các hành vi bất thường liên quan đến các mã độc lặp lại hoặc có tính chất tương tự với các mã độc đã được lưu trữ và phân tích, ở đây được gọi là cấp độ thứ nhất của việc phát hiện mã độc dựa trên các hành vi bất thường liên quan đến mã độc. Để đảm bảo rằng, việc phát hiện mã độc thông qua các hành vi bất thường trong sự kiện mà không thể được phát hiện được bằng việc phát hiện bất thường theo cấp độ thứ nhất, vẫn được tiếp tục phân tích để phát hiện tiếp, sáng chế đề xuất thêm cấp độ thứ hai của việc phát hiện mã độc. Tức là, nếu kết quả là không có bất thường trong sự

kiện gắn với tệp tin thì tiếp tục thực hiện việc phát hiện mã độc theo cấp độ thứ hai, trong đó việc phát hiện mã độc theo cấp độ thứ hai có thể được thực hiện bằng cách trích xuất các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình (Event process), các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký (Registry Event), các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng (Event Network connection), và các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin (Event File Delete). Các đặc trưng được trích xuất nêu trên sẽ được tạo thành các vec tơ đặc trưng độc lập theo từng loại sự kiện và được phân loại thông qua các mô hình học máy độc lập tương ứng.

Các đặc trưng về tiến trình được lấy làm đầu vào cho mô hình học máy để phát hiện các tiến trình bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên (Random Forest) đã được huấn luyện và kiểm thử.

Các đặc trưng về đăng ký được lấy làm đầu vào cho mô hình học máy để phát hiện các đăng ký bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Các đặc trưng về kết nối mạng được lấy làm đầu vào cho mô hình học máy để phát hiện các kết nối mạng bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Các đặc trưng về xoá tệp tin được lấy làm đầu vào cho mô hình học máy để phát hiện các tiến trình xoá tệp tin bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Tiếp theo, việc phát hiện mã độc theo cấp độ thứ hai có thể được thực hiện thông qua các bước sau:

tổng hợp ít nhất là các tiến trình bất thường, các đăng ký bất thường, các kết nối mạng bất thường, và các tiến trình xoá tệp tin bất thường đã nêu thành dạng đồ thị biểu diễn mối quan hệ sự kiện theo thời gian giữa chúng,

trích xuất, sử dụng mạng đồ thị học sâu (ví dụ mạng đồ thị học sâu GIN -

Graph Isomorphism Network), các đặc trưng về sự kiện theo thời gian gắn với tệp tin dựa trên đồ thị biểu diễn mối quan hệ sự kiện theo thời gian đã nêu,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về sự kiện theo thời gian đã nêu để phát hiện mã độc theo cấp độ thứ hai xem liệu có mã độc liên quan đến tệp tin tương ứng hay không, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Chi tiết hơn về việc phát hiện hành vi bất thường theo cấp độ thứ hai sử dụng mô hình học máy được thể hiện thông qua các bước từ bước 200 đến bước 1100 như được mô tả dưới đây.

Bước 300: tiền xử lý dữ liệu.

Ở bước này, các sự kiện sẽ được thu thập trong nhân hệ điều hành sau khi được phát hiện và phân tích tại bước 100 sẽ được chia thành các thành phần khác nhau như sự kiện liên quan đến tiến trình (Event process), sự kiện liên quan đến đăng ký (Registry Event), sự kiện liên quan đến kết nối mạng (Event Network connection), sự kiện liên quan đến xóa tệp tin (Event File Delete). Các sự kiện được sinh ra trong các tệp tin sẽ được tiền xử lý để tạo thành các dạng dữ liệu thích hợp cho việc thực hiện các bước tiếp theo.

Bước 400: trích xuất đặc trưng sự kiện liên quan đến tiến trình.

Ở bước này, dữ liệu đầu vào là các dữ liệu có trong sự kiện liên quan đến tiến trình, được trích xuất đặc trưng là các đặc trưng được xác định trước có trong sự kiện quan đến tiến trình hay còn gọi là các đặc trưng về tiến trình.

Theo một phương án ưu tiên, các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình, bao gồm nhưng không giới hạn ở, loại tiến trình; loại tệp tin; số lượng tệp mà tiến trình tạo hoặc ghi đè; và tiến trình có mức ưu tiên truy cập thấp hay không.

Bước 500: trích xuất đặc trưng sự kiện liên quan đến đăng ký.

Ở bước này, dữ liệu đầu vào là dữ liệu có sự kiện ký liên quan đến đăng ký, được trích xuất đặc trưng đăng ký là các đặc trưng được xác định trước có trong

sự kiện liên quan đến đăng ký hay còn gọi là các đặc trưng về đăng ký.

Theo một phương án ưu tiên, các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký, bao gồm nhưng không giới hạn ở, số lượng đăng ký sử dụng; số lượng HKEY_LOCAL_MACHINE; số lượng HKEY_CLASSES_ROOT; có sử dụng các khoá đăng ký khởi tạo hay không; và có sử dụng các khoá đăng ký cài đặt đang hoạt động hay không.

Bước 600: trích xuất đặc trưng của sự kiện liên quan đến kết nối mạng

Ở bước này, dữ liệu đầu vào là dữ liệu có trong sự kiện quan đến kết nối mạng, được trích xuất đặc trưng kết nối mạng là các đặc trưng được xác định trước có trong sự kiện liên quan đến kết nối mạng hay còn gọi là các đặc trưng về kết nối mạng.

Theo một phương án ưu tiên, các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng, bao gồm nhưng không giới hạn ở, tiến trình có kết nối mạng hay không; số lần kết nối mạng của tiến trình; địa chỉ IP, tên miền hoặc yêu cầu HTTP mà tiến trình kết nối tới; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không rõ nguồn gốc; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm có dấu hiệu nghi ngờ; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không hợp pháp; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm đáng tin cậy; và số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không an toàn.

Bước 700: trích xuất đặc trưng của sự kiện liên quan đến xóa tệp tin

Ở bước này, dữ liệu đầu vào là dữ liệu có trong sự kiện liên quan đến xóa tệp tin, được trích xuất đặc trưng xóa tệp tin là các đặc trưng được xác định trước có trong sự kiện liên quan đến xóa tệp tin hay còn gọi là các đặc trưng về xoá tệp tin.

Theo một phương án ưu tiên, các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin, bao gồm nhưng không giới hạn ở, số lượng các tệp tin tiến trình xóa; tổng dung lượng các tệp tin tiến trình xóa; có tiến

trình xóa tệp tin ở ổ người dùng hay không; có tiến trình xóa tệp tin ở ổ $RECYCLE.BIN hay không; có tiến trình xóa tệp tin ở ổ chứa hệ điều hành hay không; có tiến trình xóa tệp tin ở ổ chứa dữ liệu hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình (x86) hay không; có tiến trình xóa tệp tin ở ổ chứa các Perflog hay không; có tiến trình xóa tệp tin ở ổ chứa thư mục OneDriveTemp hay không; có tiến trình xóa tệp tin ở ổ chứa $WinREAgent hay không; số lượng tệp tin xml tiến trình xóa; số lượng tệp tin ảnh tiến trình xóa; số lượng tệp tin HTML tiến trình xóa; số lượng tệp tin văn bản tiến trình xóa; và số lượng tệp tin thực thi tiến trình xóa.

Tuỳ ý, phương pháp theo sáng chế có thể còn bao gồm phân loại các sự kiện dựa trên các đặc trưng được xác định trước có trong tệp thực thi, và tập hợp các đặc trưng được xác định trước có trong tệp thực thi này vào đồ thị biểu diễn mối quan hệ sự kiện theo thời đã nêu.

Bước 800: phát hiện hành vi bất thường của sự kiện theo đặc trưng.

Các đặc trưng được trích xuất nêu trên sẽ được tạo thành các vec tơ đặc trưng độc lập theo từng loại sự kiện và được phân loại thông qua các mô hình học máy độc lập tương ứng.

Các đặc trưng về tiến trình được lấy làm đầu vào cho mô hình học máy để phát hiện các tiến trình bất thường hay các hành vi bất thường của sự kiện liên quan đến tiến trình.

Các đặc trưng về đăng ký được lấy làm đầu vào cho mô hình học máy để phát hiện các đăng ký bất thường hay các hành vi bất thường của sự kiện liên quan đến đăng ký.

Các đặc trưng về kết nối mạng được lấy làm đầu vào cho mô hình học máy để phát hiện các kết nối mạng bất thường hay các hành vi bất thường của sự kiện liên quan đến kết nối mạng.

Các đặc trưng về xoá tệp tin được lấy làm đầu vào cho mô hình học máy để

phát hiện các tiến trình xoá tệp tin bất thường hay các hành vi bất thường của sự kiện liên quan đến xoá tệp tin.

Ưu tiên là, các mô hình học máy được sử dụng là các mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Bước 900: xây dựng hồ sơ sự kiện.

Ở bước này các sự kiện kèm theo các hành vi bất thường liên quan đến sự kiện được phát hiện theo đặc trưng đã nêu sẽ được tổng hợp và kết nối với nhau thông qua việc xây dựng hồ sơ sự kiện. Cụ thể hơn, các sự kiện này sẽ được kết nối và xây dựng thành một đồ thị. Nguyên tắc xây dựng đồ thị là tất cả các sự kiện được sinh ra trên cùng một tệp tin, ví dụ tệp tin thực thi thì sẽ được gom cùng nhau dựa trên các định danh (ID) của các sự kiện. Các đỉnh của đồ thị là các sự kiện và cạnh của đồ thị là việc một sự kiện cha gọi sự kiện con con. Mỗi đỉnh trong đồ thị chứa thuộc tính là nhãn đã được phân loại, ví dụ có hành vi bất thường của sự kiện liên quan đến tiến trình, hành vi bất thường của sự kiện liên quan đến đăng ký, hành vi bất thường của sự kiện liên quan đến kết nối mạng, hành vi bất thường của sự kiện liên quan đến xoá tệp tin.

Bước 1000: trích xuất đặc trưng của mã độc dựa trên hồ sơ sự kiện

Ở bước này, dữ liệu đầu vào là dữ liệu dạng đồ thị nêu trên, mà biểu diễn mối quan hệ giữa đỉnh và cạnh của các sự kiện. Đồ thị này sẽ được chuẩn hóa và biến đổi thành một vec tơ đặc trưng của mã độc duy nhất, có chiều dài cố định. Chiều dài của vec tơ này có thể là 128, 256, hoặc 512 chiều. Điều này có thể được thực hiện thông qua sử dụng mạng đồ thị học sâu GIN (Graph Isomorphism Network) sẽ được mô tả chi tiết hơn dưới đây .

Bước 1100: phát hiện hành vi bất thường của mã độc theo đặc trưng.

Ở bước này hành vi bất thường của mã độc được phát hiện theo đặc trưng. Cụ thể hơn, vec tơ đặc trưng của mã độc được sử dụng làm đầu vào của mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử để phân loại, đưa ra kết quả liệu có mã độc thông qua sự bất thường gắn với các sự kiện của tệp tin, ví dụ tệp tin

thực thi hay không.

Như vậy, việc phát hiện hành vi bất thường của mã độc theo sáng chế được thực hiện theo hai cấp độ, không chỉ phát hiện nhanh chóng và chính xác các hành vi bất thường của mã độc hoặc mã độc đối với các mã độc lặp lại hoặc có tính chất tương tự với các mã độc đã được lưu trữ và phân tích, mà còn có khả năng phát hiện sớm các hành vi bất thường của mã độc, có tính đến nhiều đặc trưng đa dạng khác nhau một cách tổng thể, để hỗ trợ hiệu quả cho việc giám sát dữ liệu thực tế trong các hệ thống phát hiện mã độc.

Người có hiểu biết trung bình về lĩnh vực kỹ thuật này hiểu rằng, các đặc trưng của mã độc được trích xuất từ sự kiện trong các khoảng thời gian thích hợp, ví dụ trong khoảng thời gian một giờ, một ngày, một tuần, một tháng, nhưng không chỉ giới hạn ở đó mà có thể là bất kỳ một khoảng thời gian được xác định trước nào đó thích hợp với đặc tính hoạt động tấn công bằng mã độc chẳng hạn.

Có thể thấy là, phương pháp theo sáng chế tập trung vào phân tích và đánh giá các sự kiện được sinh ra từ tệp tin thực thi rồi từ đó tiến hành trích xuất các đặc trưng mà có thể phản ánh các hành vi bất thường từ các của sự kiện này, và hướng tới việc thu thập được tối đa các đặc trưng khác nhau thể hiện sự tương quan giữa các sự kiện để làm cơ sở cho việc dự báo hoặc phân loại các hành vi bất thường, từ đó nâng cao khả năng phát hiện được hành vi bất thường của mã độc. Việc xây dựng hồ sơ đặc trưng sự kiện trong tệp tin thông qua việc biểu diễn mối quan hệ giữa các sự kiện: tiến trình, đăng ký, kết nối mạng và xóa tệp tin rồi trích xuất đặc trưng của hồ sơ đặc trưng của sự kiện dạng đồ thị thành thành một vec tơ đặc trưng duy nhất có thể được hiểu như là việc xây dựng hồ sơ hành vi bất thường bao hàm cơ bản là các hành vi bất thường có thể xuất hiện tệp tin thực thi dựa trên chính các thành phần khác nhau trong sự kiện do tệp tin thực thi sinh ra. Nhờ đó khắc phục được nhược điểm của các phương pháp phát hiện hành vi bất thường của tệp tin thực thi chỉ dựa trên các đặc trưng được trích xuất từ một hoặc một số thành phần nào đó của tệp tin thực thi. Với cách thức như vậy, sáng chế được cho là có thể tổng hợp được đầy đủ thông tin quan trọng và có ý nghĩa được

chứa trong sự kiện được sinh ra từ tệp tin thực thi, và do đó có thể phản ánh được sự tương quan vốn có của các sự kiện khác nhau của mã độc, được cho là có thể phát hiện được các mối nguy hiểm mới hoặc hành vi bất thường mới liên quan đến các hành vi bất thường của tệp tin thực thi.

Dễ thấy là, sau khi xây dựng hồ sơ sự kiện theo như bước 900 nêu trên, có thể sử dụng trực tiếp hồ sơ sự kiện này để phân tích nhằm phát hiện mã độc, mà không sử dụng đồ thị. Tuy nhiên, điều này có vấn đề liên quan đến sự khác nhau về số lượng sự kiện liên quan đến tệp tin, hay số lượng sự kiện trong từng tệp tin thường sẽ khác nhau. Có tệp tin thì có tới vài chục, vài trăm thậm chí vài chục nghìn sự kiện. Tuy nhiên, cũng có tệp tien thì chỉ có một hoặc vài sự kiện. Dẫn đến vấn đề cần đặt ra là làm thế nào để chuẩn hóa được số lượng sự kiện trong các tệp tin để kết luận hành vi bất thường liên quan đến mã độc. Nếu chỉ dựa vào một sự kiện bất thường gắn với tệp tin để kết luận luôn đây là mã độc sẽ có thể dẫn đến cảnh báo sai rất nhiều, không hiệu quả thực tế. Nếu chuẩn hóa số lượng sự kiện này, Ví dụ tệp tin thứ nhất có 100 sự kiện, tệp tin thứ hai có 200 sự kiện, việc chuẩn hóa đặt ra là lấy mỗi tệp tin 100 sự kiện để phân tích, sẽ có thể dẫn đến bỏ sót các sự kiện quan trong do không thể biết được sự kiện nào quan trọng sự kiện nào không quan trọng khi chuẩn hóa. Ngoài ra, đối với các tệp tin cụ thể sẽ có thể không xác định được đến khi nào thì xuất hiện đủ số lượng các sự kiện, giả sử tệp tin nào đó có chứa mã độc thì khó thực hiện việc phát hiện sớm nếu như số lượng sự kiện gắn với tệp tin này là ít, chẳng hạn.

Hơn nữa, do các sự kiện được tập hợp thành từng nhóm được sinh ra theo tệp tin không biểu thị được mối quan hệ giữa các tiến trình nên sẽ khó nhận biết được các hành vi bất thường của mã độc dựa trên sự kiện. Thực tế cho thấy, mã độc sẽ cố gắng sử dụng các sự kiện bình thường để che giấu các hành vi của chúng. Tuy nhiên, nếu các sự kiện này được ghép nối với nhau thì có thể nhận thấy được hành vi bất thường của mã độc. Bởi vì trong thực tế là hành vi của mã độc không chỉ thể hiện ở từng tiến trình đơn lẻ mà phải mà còn là sự kết hợp các sự kiện. Ví dụ, hôm nay người dùng nhận được một thư điện tử, rồi tải tệp tin có định dạng Docx về máy tính. Sau đó, khoảng 3 tiếng sau thì người dùng mở tệp tin docx đó

lên. Khoảng 24 giờ sau thì máy tính kết nối đến 1 địa chỉ IP. Hệ thống giám sát và phát hiện mã độc nhận thấy được IP mà máy tính kết nối đến hoàn toàn bình thường. Như vậy, có thể thấy rằng, nếu dựa trên các sự kiện riêng lẻ thì hệ thống không nhận thấy được bất kỳ sự kiện nào vì mã độc đã che giấu rất tốt cũng như thực hiện kết nối hoàn toàn hợp lệ. Tuy nhiên, nếu các sự kiện này được kết hợp với nhau thành một đồ thị thì hệ thống sẽ thấy được mối quan hệ giữa các sự kiện này. Mối quan hệ giữa các sự kiện này đó là: tải tệp tin tài liệu → mở tệp tin tài liệu → kết nối với địa chỉ IP bên ngoài. Từ đồ thị này dễ dàng nhận thấy ngay được đây là hành vi tấn công thông qua mã độc vì không có sự kiện mở tệp tin nào mà lại mở kết nối đến địa chỉ bên ngoài. Như vậy có thể thấy rằng nếu không có việc kết hợp dạng đồ thị này thì không thể biết được hành vi bất thường của nhiều dạng mã độc.

Như vậy, việc xây dựng thành dạng đồ thị được cho là đặc trưng khác biệt có thể mang lại hiệu quả tốt cho sáng chế.

Hình 2 là giản đồ giản lược thể hiện việc huấn luyện và sử dụng mô hình học máy trong việc phát hiện bất thường trong sự kiện theo sáng chế.

Theo một phương án ưu tiên, mô hình học máy theo sáng chế là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

Nói chung, mô hình rừng ngẫu nhiên đã được đề xuất sử dụng trong một số giải pháp phát hiện bất thường trong mạng liên quan đến các kịch bản tấn công nào đó, chẳng hạn như trong bài báo có tiêu đề là "Detection System of HTTP DDoS Attacks in a Cloud Environment Based on Information Theoretic Entropy and Random Forest", được công bố ngày 5/6/2018 của tác giả Mohamed Idhammad và các cộng sự. Do đó, các mô tả chi tiết về việc ứng dụng mô hình rừng ngẫu nhiên được dự định lược bỏ để tránh các thông tin dư thừa. Dưới đây chỉ mô tả các bước tổng quát của việc huấn luyện và sử dụng mô hình học máy trong việc phát hiện bất thường trong sự kiện theo sáng chế.

Trong phần (a) thể hiện giai đoạn huấn luyện mô hình học máy. Các sự kiện được các tệp tin sinh ra sẽ được trích xuất các đặc trưng để làm tập dữ liệu sử

dụng trong quá trình huấn luyện mô hình học máy. Thông thường các đặc trưng được trích xuất được gán nhãn và được chia thành tập dữ liệu huấn luyện để huấn luyện mô hình học máy và tập dữ liệu kiểm tra để kiểm thử mô hình học máy được huấn luyện.

Trong phần (b) thể hiện giai đoạn sử dụng mô hình học máy. Các sự kiện được trích xuất các đặc trưng để làm dữ liệu đầu vào cho việc phân loại sử dụng mô hình học máy đã được huấn luyện và kiểm thử. Kết quả phân loại có thể là bình thường hoặc bất thường.

Tiếp theo, về thuật toán thể hiện nguyên tắc xây dựng hồ sơ sự kiện của tệp tin thực thi được sinh ra từ tệp tin thực thi theo một phương án ưu tiên của sáng chế sẽ được trích dẫn để làm ví dụ.

"Thuật toán: xây dựng đồ thị biểu diễn mối quan hệ của các sự kiện của tệp tin

Input : Danh sách sự kiện đã được gán nhãn
Output : Đồ thị
A: ma trận kề
X: ma trận thuộc tính
labeled_processes : danh sách sự kiện đã được gán nhãn của file
n = lenth(labeled_processes)
A = init maxtrix size nxn with fill value = 0
X = init array size 1xn with fill value = -1
each process in labeled_processes:
parent_node , curr_node, feature = get_graph_info(process)
A= update(A, parent_node , curr_node)
X = update(X ,curr_node, feature)
return Graph(A, X)".

Trên Hình 3 thể hiện giản đồ giản lược ví dụ về hồ sơ hành vi mã độc thể hiện mối quan hệ giữa các sự kiện dạng cây được sinh ra dựa trên thuật toán được trích dẫn trên đây.

Hình 4 là giản đồ giản lược thể hiện việc huấn luyện và sử dụng mô hình học máy trong việc phát hiện mã độc theo sáng chế.

Tương tự như các nội mô tả gắn với Hình 2, mô hình học máy được minh hoạ ở đây cũng được lựa chọn là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử để phát hiện phát hiện hành vi bất thường của mã độc theo đặc trưng được trích xuất từ đồ thị, và được phân chia thành hai phần (a) và (b) như được mô tả chi tiết hơn dưới đây.

Trong phần (a) thể hiện giai đoạn huấn luyện mô hình học máy. Các hồ sơ sự kiện được xây dựng dựa trên các sự kiện trong tệp tin, ví dụ tệp tin thực thi, được trích xuất các đặc trưng để làm tập dữ liệu sử dụng trong quá trình huấn luyện mô hình học máy. Thông thường các đặc trưng được trích xuất được gán nhãn và được chia thành tập dữ liệu huấn luyện để huấn luyện mô hình học máy và tập dữ liệu kiểm tra để kiểm thử mô hình học máy được huấn luyện.

Trong phần (b) thể hiện giai đoạn sử dụng mô hình học máy. Các hồ sơ sự kiện được xây dựng dựa trên các sự kiện trong tệp tin, ví dụ tệp tin thực thi được trích xuất các đặc trưng để làm dữ liệu đầu vào cho việc phân loại sử dụng mô hình học máy đã được huấn luyện và kiểm thử. Kết quả phân loại có thể là hành vi bình thường của tệp tin thực thi hoặc hành vi bất thường của tệp tin thực thi.

Tiếp theo sẽ mô tả một số hướng dẫn là ví dụ minh hoạ cho việc thực hiện sáng chế. Hướng dẫn này cơ bản là được thể hiện theo các bước từ bước 100 đến bước 1100 như được mô tả trên đây, tuy nhiên, sáng chế không bị giới hạn ở đó.

Trước tiên, bước 100 được thực hiện theo một số nội dung dưới đây.

Nội dung thứ nhất: cài đặt và tạo cấu hình công cụ thu thập sự kiện trên máy người dùng bằng công cụ Wahzu.

Công cụ thu thập sự kiện trên máy người dùng Wahzu (hay có thể gọi tắt là Wahzu) là một ứng dụng mã nguồn mở đa nền tảng. Wahzu có thể được cài đặt bằng nhiều cách theo hướng dẫn từ tài liệu chính thức. Ví dụ sử dụng cài đặt gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu theo các lệnh sau:

Cài đặt gói đã tải xuống bằng lệnh dpkg:

sudo dpkg -i wazuh-agent_4.2.1-1_amd64.deb

Mở tệp cấu hình của Wazuh agent bằng trình soạn thảo văn bản:

sudo nano/var/ossec/etc/ossec.conf

Chỉnh sửa địa chỉ IP hoặc tên miền của máy chủ Wazuh và mật khẩu của máy chủ Wazuh trong tệp cấu hình. Sau đó, lưu và đóng tệp. Khởi động lại dịch vụ Wazuh agent để bắt đầu gửi thông tin về máy chủ Wazuh:

sudo systemctl restart wazuh-agent.service

Sau khi hoàn thành các bước trên, máy tính Ubuntu sẽ được theo dõi và báo cáo các hoạt động của nó đến máy chủ Wazuh. Có thể kiểm tra trạng thái của Wazuh agent bằng lệnh sau:

sudo systemctl status wazuh-agent.service Nếu trạng thái hiển thị là "active (running)", thì Wazuh agent đang hoạt động bình thường và gửi dữ liệu đến máy chủ Wazuh.

Nội dung thứ 2: thu thập sự kiện. Sau khi cài đặt thành công công cụ Wahzu trên máy tính người dùng. Công cụ này sẽ tiến hành thu thập các tiến trình trên máy người dùng. Công cụ có thể thu thập được nhiều tiến trình khác nhau. Trong các tài liệu mô tả về công cụ Wahzu đã liệt kê chi tiết các sự kiện này.

Tiếp theo, bước 200 được thực hiện theo một số nội dung dưới đây.

Nội dung thứ nhất: cài đặt và cấu hình công cụ lưu trữ Wahzu. Gọi tắt là máy chủ Wahzu. Công cụ này có chức năng thu nhận các sự kiện được gửi lên từ công cụ Wahzu trên máy người dùng. Máy chủ Wahzu có thể được cài đặt bằng nhiều cách theo hướng dẫn từ tài liệu chính thức. Ví dụ sử dụng cài đặt gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu.

Cài đặt các gói cần thiết trong Ubuntu bằng lệnh sau:

sudo apt-get update sudo apt-get install curl apt-transport-https

lsb-release gnupg2

Thêm kho lưu trữ Wazuh vào hệ thống bằng lệnh sau:

curl -s https://packages.wazuh.com/key/GPG-KEY-WAZUH | sudo apt-key add -

echo "deb https://packages.wazuh.com/4.x/apt/ stable main" | sudo tee /etc/apt/sources.list.d/wazuh.list

Cập nhật lại danh sách các gói và cài đặt Wazuh server bằng lệnh sau:

sudo apt-get update

sudo apt-get install wazuh-manager

Sau khi cài đặt xong, có thể truy cập giao diện web của Wazuh server bằng cách mở trình duyệt và truy cập địa chỉ IP của máy chủ theo định dạng sau: https://<địa chỉ IP của máy chủ>:55000.

Đăng nhập vào giao diện web bằng tài khoản admin mặc định bằng cách sử dụng thông tin đăng nhập sau:

Tên đăng nhập: admin

Mật khẩu: admin

Nội dung 2: xây dựng tập luật cho phát hiện hành vi bất thường của mã độc.

Theo đó, tập luật để phát hiện hành vi bất thường của mã độc được xây dựng trên nền tảng của MITRE ATT&CK Framework. MITRE ATT&CK Framework là một nền tảng phát hiện các hành vi bất thường dựa trên các chiến lược và chiến thuật. Dựa trên các chiến lược và chiến thuật này, có thể xây dựng được các tập luật. Ví dụ xây dựng tập luật dựa trên MITRE ATT&CK Framework như sau: chuyển đổi "T1053.002 – Scheduled Task" thành luật.

Ví dụ, kẻ tấn công có thể lạm dụng tiện ích at.exe để thực hiện lập lịch tác vụ nhằm thực thi mã độc hại lần đầu hoặc định kỳ. Tiện ích at.exe tồn tại dưới dạng tệp thực thi trong Windows để lập lịch tác vụ tại một thời điểm và ngày cụ thể. Việc sử dụng "at" yêu cầu dịch vụ "Task Scheduler" đang chạy và người dùng phải đăng nhập với tư cách là thành viên của nhóm Quản trị (Administrators

group). Kẻ tấn công có thể sử dụng at.exe trong môi trường Windows để thực thi các chương trình khi khởi động hệ thống hoặc trên cơ sở đã lên lịch để tồn tại lâu dài. "at" cũng có thể bị lạm dụng để tiến hành Thực thi từ xa (Remote Execution) như một phần của leo thang đặc quyền (Lateral Movement) và hoặc để chạy một chương trình trong ngữ cảnh của một tài khoản được chỉ định (như tài khoản hệ thống).

Cách nhận diện: giám sát việc thực thi quy trình từ svchost.exe trong Windows 10 và Trình lập lịch tác vụ Windows (Windows Task Scheduler) taskeng.exe cho các phiên bản Windows cũ hơn. Giám sát Windows Task Scheduler lưu trữ trong %systemroot%\System32\Tasks cho các mục thay đổi liên quan đến các tác vụ đã lên lịch không tương quan với phần mềm đã biết, các chu kỳ vá lỗi, v.v.. Quá trình thực hiện:

- Chọn create new rule → event correlation.
- Tại phần index pattern, chọn index mong muốn (filebeat / winlogbeat).
- Tại phần custom query, t có câu query sau:

process where event.type == "start" and / Schedule service cmdline on Win10+ */ process.parent.name : "svchost.exe" and process.parent.args : "Schedule" and /* add suspicious programs here */ process.pe.original_file_name in ("cscript.exe", "wscript.exe", "PowerShell.EXE", "Cmd.Exe", "MSHTA.EXE", "RUNDLL32.EXE", "REGSVR32.EXE", "MSBuild.exe", "InstallUtil.exe", "RegAsm.exe", "RegSvcs.exe", "msxsl.exe", "CONTROL.EXE", "EXPLORER.EXE", "Microsoft.Workflow.Compiler.exe", "msiexec.exe") and /* add suspicious paths here */ process.args : ("C:\\Users*", "C:\\ProgramData*", "C:\\Windows\\Temp*", "C:\\Windows\\Tasks*", "C:\\PerfLogs*", "C:\\Intel*", "C:\\Windows\\Debug*", "C:\\HP*")*

Tiếp theo, ở bước 300, sau khi các sự kiện được phân tích và phát hiện dựa trên tập luật được xây dựng từ MITRE ATT&CK Framework. Hệ thống giám sát và phát hiện mã độc dựa trên tập luật sẽ thực hiện phân tích và thông báo ra nhật ký kết quả giám sát đầy đủ dưới hai dạng văn bản và Json. Để tiếp tục phát hiện

các hành vi bất thường của mã độc, sáng chế dựa trên việc phân tích bốn sự kiện chính liên quan đến sự kiện tiến trình, sự kiện đăng ký, sự kiện kết nối mạng và sự kiện xóa tệp tin theo các đặc trưng được xác định trước một cách đầy đủ như được thể hiện rõ hơn trong các nội dung tiếp theo tương ứng với các bước từ bước 400 đến bước 700.

Tiếp theo, bước 400: trích xuất đặc trưng sự kiện tiến trình.

Từ dữ liệu nhật ký của máy chủ Wahzu dưới dạng Json, người dùng có thể trích xuất được những đặc trưng quan trọng của sự kiện tiến trình phản ánh các hành vi bất thường hay không, các đặc trưng này được thể hiện trong Bảng 1 dưới đây.

Bảng 1. Danh sách các đặc trưng của sự kiện tiến trình

Tên	Mô tả
ID tiến trình	Định danh của tiến trình kích hoạt sự kiện
ParentPID	Giống như ID tiến trình nhưng dành cho tiến trình gốc
Creation Times tamp	Thời gian khi tiến trình được tạo
Chuỗi dòng lệnh	Chuỗi Dòng lệnh chứa tất cả các đối số được truyền cho tiến trình khi thực thi
Hình ảnh	Hình ảnh Đường dẫn tệp của tệp thực thi được liên kết với quy trình
Process Type	Loại tiến trình
File Type	Loại tệp tin
Tệp tin (Files)	Số lượng file mà tiến trình tạo hoặc ghi đè.
Modules	Số lượng mô đun mà tiến trình sử dụng
Low Access	Tiến trình có mức ưu tiên truy cập thấp hay không?
Autostart	Tiến trình có khả năng tự khởi động không?
Priority	Mức độ ưu tiên xác định thứ tự của tiến trình được thực hiện so với các tiến trình khác cần được hoàn thành. Có 3 mức độ là cao, trung bình và thấp.

Tiếp theo, bước 500: trích xuất đặc trưng sự kiện đăng ký.

Từ dữ liệu nhật ký của máy chủ Wahzu dưới dạng Json, người dùng có thể

trích xuất được những đặc trưng quan trọng của sự kiện đăng ký phản ánh các hành vi bất thường hay không, các đặc trưng này được thể hiện trong Bảng 2 dưới đây.

Bảng 2. Danh sách các đặc trưng của sự kiện đăng ký

Tên	Mô tả
Đăng ký (Registry)*	Số lượng registry sử dụng.
HKEY_LOCAL_MACHINE*	Số lượng HKEY_LOCAL_MACHINE
HKEY_CLASSES_ROOT*	Số lượng HKEY_CLASSES_ROOT
HKEY_USERS*	Số lượng HKEY_USERS
HKEY_CURRENT_CONFIG*	Số lượng HKEY_CURRENT_CONFIG
HKEY_CURRENT_USER*	Số lượng HKEY_CURRENT_USER
Khóa đăng ký khởi động	Có sử dụng hay không?
Khóa đăng ký thiết lập đang hoạt động	Có sử dụng hay không?
Khóa đăng ký dịch vụ	Có sử dụng hay không?
Khóa đăng ký DLL	Có sử dụng hay không?
Vỏ bọc sinh ra các khóa đăng ký	Có sử dụng hay không?
Khóa đăng ký cài đặt internet	Có sử dụng hay không?
Khóa đăng ký BHO	Có sử dụng hay không?

Tiếp theo, bước 600: trích xuất đặc trưng kết nối mạng.

Từ dữ liệu nhật ký của máy chủ Wahzu dưới dạng Json, người dùng có thể trích xuất được những đặc trưng quan trọng của sự kiện kết nối mạng phản ánh các hành vi bất thường hay không, các đặc trưng này được thể hiện trong Bảng 3 dưới đây.

Bảng 3. Danh sách các đặc trưng của sự kiện kết nối mạng

Tên	Kiểu	Mô tả
Điểm mạng	Binary	Tiến trình có kết nối mạng hay không?
Đếm sự kiện của mạng	Number	Số lần kết nối mạng của tiến trình.
Địa chỉ IP/tên miền/Yêu cầu http		Địa chỉ ip, domain hoặc HTTP Request mà tiến trình kết nối tới.

Địa chỉ IP không xác định/tên miền không xác định /yêu cầu không xác định	Number	Số lượng các địa chỉ IP, tên miền, yêu cầu thuộc nhóm không rõ nguồn gốc
Địa chỉ IP đáng ngờ /tên miền không xác định / yêu cầu không xác định	Number	Số lượng các địa chỉ IP, tên miền, yêu cầu thuộc nhóm có dấu hiệu nghi ngờ
Địa chỉ IP độc hại/Tên miền độc hại/ yêu cầu độc hại	Number	Số lượng các địa chỉ IP, tên miền, yêu cầu thuộc nhóm không hợp pháp
Địa chỉ IP bình thường/tên miền bình thường/ yêu cầu bình thường	Number	Số lượng các địa chỉ IP, tên miền, yêu cầu thuộc nhóm đáng tin cậy
Địa chỉ IP không an toàn/tên miền không an toàn/ yêu cầu không an toàn	Number	Số lượng các địa chỉ IP, tên miền, yêu cầu thuộc nhóm không an toàn
Trọng đố của địa chỉ IP	Number	Trọng số của của IP được tính theo công thức sau: trọng số = **(không xác định* 0 + nghi ngờ * 1 + độc hại * 2 + bình thường* 3 + không an toàn * 4) / 5**
Trọng số của tên miền	Number	Trọng số của của Domain được tính theo công thức sau: trọng số = **(không xác định* 0 + nghi ngờ * 1 + độc hại * 2 + bình thường* 3 + không an toàn * 4)/ 5**
Trọng số của yêu cầu	Number	Trọng số của của Request được tính theo công thức sau: Trọng số = **(không xác định* 0 + nghi ngờ * 1 + độc hại * 2 + bình thường* 3 + không an toàn * 4)/ 5**

Tiếp theo, bước 700: trích xuất đặc trưng xóa tệp tin.

Từ dữ liệu nhật ký của máy chủ Wahzu dưới dạng Json, người dùng có thể trích xuất được những đặc trưng quan trọng của sự kiện kết nối mạng phản ánh các hành vi bất thường hay không, các đặc trưng này được thể hiện trong Bảng 4 dưới đây.

Bảng 4. Danh sách các đặc trưng tệp tin trong nhật ký Suricata

Tên	Mô tả
Drop Files	Số lượng các tệp tin tiến trình bị xóa
Total drop file size	Tổng dung lượng các tệp tin tiến trình bị xóa
Users	Tiến trình xóa tệp tin ở ổ User hay không?

$RECYCLE.BIN	Tiến trình xóa tệp tin ở ổ $RECYCLE.BIN hay không?
Windows	Tiến trình xóa tệp tin ở ổ Windows hay không?
Program Data	Tiến trình xóa tệp tin ở ổ ProgramData hay không?
Program Files	Tiến trình xóa tệp tin ở ổ Program Files hay không?
Program Files (x86)	Tiến trình xóa tệp tin ở ổ Program Files (x86) hay không?
Perflogs	Tiến trình xóa tệp tin ở ổ Perflogs hay không?
OneDriveTemp	Tiến trình xóa tệp tin ở ổ OneDriveTemp hay không?
$WinREAgent	Tiến trình xóa tệp tin ở ổ $WinREAgent hay không?
Type XML	Số lượng tệp tin định dạng xml tiến trình xóa
Type Image	Số lượng tệp tin định dạng ảnh tiến trình xóa
Type HTML	Số lượng tệp tin định dạng HTML tiến trình xóa
Type Text	Số lượng tệp tin văn bản tiến trình xóa
Type Executable	Số lượng tệp tin thực thi tiến trình xóa

Cuối cùng, bước 800: phân loại sự kiện dựa trên các hành vi bất thường.

Ở bước này hành vi bất thường của sự kiện được phát hiện theo đặc trưng. Cục thể hơn, vec tơ đặc trưng của các sự kiện được sử dụng làm đầu vào của mô hình rừng ngẫu nhiên (Random Forest) đã được huấn luyện và kiểm thử để phân loại, đưa ra kết quả liệu có bất thường trong sự kiện hay không.

Tiếp theo, ở bước 900: xây dựng hồ sơ sự kiện

Dựa trên tất cả các đặc trưng bất thường thu được từ quá trình phân tích ở trên. Sáng chế sẽ tiến hành tổng hợp lại các sự kiện trên theo từng tệp tin. Có nghĩa là bước này sẽ tổng hợp và gộp các sự kiện được sinh ra trên cùng một tệp tin thực thi theo nguyên tắc tất cả các sự kiện được sinh ra trên cùng 1 tệp tin thực thi thì sẽ được gom cùng nhau dựa trên các ID của các sự kiện. Các đỉnh của đồ

thị là các sự kiện và cạnh của đồ thị là việc một sự kiện cha gọi sự kiện con con. Mỗi đỉnh trong đồ thị chứa thuộc tính là nhãn vừa được phân loại.

Tiếp theo, bước 1000: trích xuất đặc trưng mã độc dựa trên hồ sơ tiến trình.

Dữ liệu đầu vào của bước này là dữ liệu dạng đồ thị đã được xây dựng ở trên, sáng chế sẽ tiến hành biến đổi và chuyển hóa dữ liệu đồ thị với các kích thước khác nhau này về dạng vec tơ đặc trưng, duy nhất và có chiều dài bằng nhau thông qua mô hình mạng đồ thị học sâu GIN (Graph Isomorphism Network). Mạng GIN là một GNN (Graph Neural Network) dạng SGCN (Spatial Graph Convolutional Network). Hiện nay, GIN đang được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau như: dự đoán ADMET (absorption, distribution, metabolism, exclusion, and toxicity) trong thực phẩm, mĩ phẩm, dược phẩm, phân tích dữ liệu ảnh y khoa, v.v..

Nói chung mô hình GIN đã được đề xuất và mô tả chi tiết trong bài báo có tiêu đề là "HOW POWERFUL ARE GRAPH NEURAL NETWORKS", được công bố ngày 22/02/2019 của tác giả Keyulu Xu và các cộng sự. Do đó, các mô tả chi tiết về việc ứng dụng mô hình GIN được dự định lược bỏ để tránh các thông tin dư thừa. Dưới đây chỉ mô tả các bước tổng quát của việc sử dụng mô hình GIN trong việc trích xuất đặc trưng tệp tin thực thi theo sáng chế.

Nguyên tắc trích xuất đặc trưng mã độc trong hồ sơ sự kiện sử dụng mạng GIN như sau: với đầu vào là hồ sơ sự kiện, mạng GIN sẽ thực hiện trích xuất đặc điểm của đồ thị liên kết giữa các sự kiện đầu ra của các lớp GIN là các biểu diễn đặc trưng ẩn (hidden features representation) thể hiện sự kết nối giữa các nút trong đồ thị. Trong đó, lớp GIN thứ nhất có nhiệm vụ xem xét các đỉnh lân cận trên từng đỉnh để cập nhật lại thuộc tính của đỉnh đó. Các lớp GIN tiếp theo có chức năng cập nhật đệ quy thuộc tính của các nút trong đồ thị.

Cuối cùng, bước 1100: phát hiện hành vi bất thường của mã độc theo đặc trưng.

Ở bước này hành vi bất thường của mã độc được phát hiện theo đặc trưng. Cục thể hơn, vec tơ đặc trưng của tệp tin thực thi được sử dụng làm đầu vào của

mô hình rừng ngẫu nhiên (Random Forest) đã được huấn luyện và kiểm thử để phân loại, đưa ra kết quả liệu có bất thường trong tệp tin, ví dụ tệp tin thực thi hay không.

Theo một khía cạnh khác, sáng chế đề xuất phương tiện lưu trữ có chứa chương trình máy tính bao gồm các lệnh mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp phát hiện bất thường trong tệp tin, ví dụ tệp tin thực thi trên đây. Phương tiện lưu trữ này có thể là một hoặc sự kết hợp của nhiều phương tiện hoặc bất kỳ, có thể kết nối theo cách trực tiếp, gián tiếp, từ xa, hoặc qua điện toán đám mây, nhưng không chỉ giới hạn ở đó.

Trên đây, sáng chế đã được mô tả chi tiết theo các phương án ưu tiên thực hiện, và có thể kèm theo các phương án thay thế hoặc tương đương hoặc các ví dụ cụ thể, sử dụng các mô tả và thuật ngữ phù hợp để người có hiểu biết trung bình về lĩnh vực kỹ thuật này có thể hiểu và thực hiện được giải pháp theo sáng chế. Vì vậy, người có hiểu biết trung bình về lĩnh vực kỹ thuật tương ứng có thể dễ dàng tạo ra các thay đổi, cải biến, hoặc thay thế tương đương dựa vào các nội dung và phương án được mô tả. Do đó, các thay đổi, cải biến, hoặc thay thế tương đương này được coi là không nằm ngoài phạm vi của sáng chế, và phạm vi bảo hộ của sáng chế hiển nhiên là không bị giới hạn bởi các nội dung và phương án được mô tả mà được xác định trong yêu cầu bảo hộ dưới đây.

Yêu cầu bảo hộ

1. Phương pháp phát hiện mã độc bao gồm:

thu thập, bởi công cụ phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng, các dấu hiệu có trong các sự kiện gắn với tệp tin, các sự kiện gắn với tệp tin này bao gồm ít nhất là sự kiện tiến trình (Event process), sự kiện đăng ký (Registry Event), sự kiện kết nối mạng (Event Network connection), sự kiện xóa tệp tin (Event File Delete);

so khớp các dấu hiệu trong các sự kiện đã nêu với tập dấu hiệu mã độc hoặc kiểm tra các dấu hiệu trong các sự kiện gắn với tệp tin đã nêu bằng tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc, để thực hiện việc phát hiện mã độc theo cấp độ thứ nhất xem liệu có mã độc liên quan tới tệp tin tương ứng hay không,

trong đó tập dấu hiệu mã độc và tập luật phát hiện mã độc dựa trên các dấu hiệu về hành vi bất thường của mã độc đã có trong cơ sở dữ liệu, được tổng hợp từ các mã độc đã biết trước đó, và tập luật phát hiện mã độc được xây dựng nhờ sử dụng MITRE ATT&CK (Adversarial Tactics Techniques & Common Knowledge) Framework;

nếu kết quả là không có mã độc thì tiếp tục thực hiện việc phát hiện mã độc theo cấp độ thứ hai, trong đó việc phát hiện mã độc theo cấp độ thứ hai được thực hiện thông qua các bước sau:

trích xuất các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình (Event process), các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký (Registry Event), các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng (Event Network connection), và các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin (Event File Delete),

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về tiến trình đã nêu để phát hiện các tiến trình bất thường, trong đó mô hình học máy là

mô hình rừng ngẫu nhiên (Random Forest) đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về đăng ký đã nêu để phát hiện các đăng ký bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về kết nối mạng đã nêu để phát hiện các kết nối mạng bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về xoá tệp tin đã nêu để phát hiện các tiến trình xoá tệp tin bất thường, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử,

tổng hợp ít nhất là các tiến trình bất thường, các đăng ký bất thường, các kết nối mạng bất thường, và các tiến trình xoá tệp tin bất thường đã nêu thành dạng đồ thị biểu diễn mối quan hệ sự kiện theo thời gian giữa chúng,

trích xuất, sử dụng mạng đồ thị học sâu, các đặc trưng về sự kiện theo thời gian gắn với tệp tin dựa trên đồ thị biểu diễn mối quan hệ sự kiện theo thời gian đã nêu,

phân loại, bởi mô hình học máy, dựa trên đầu vào là các đặc trưng về sự kiện theo thời gian đã nêu để phát hiện mã độc theo cấp độ thứ hai xem liệu có mã độc liên quan đến tệp tin tương ứng hay không, trong đó mô hình học máy là mô hình rừng ngẫu nhiên đã được huấn luyện và kiểm thử.

2. Phương pháp theo điểm 1, trong đó các đặc trưng về tiến trình được xác định trước có liên quan đến sự kiện tiến trình, bao gồm nhưng không giới hạn ở, loại tiến trình; loại tệp tin; số lượng tệp mà tiến trình tạo hoặc ghi đè; và tiến trình có mức ưu tiên truy cập thấp hay không.

3. Phương pháp theo điểm 1 hoặc 2, trong đó các đặc trưng về đăng ký được xác định trước có liên quan đến sự kiện đăng ký, bao gồm nhưng không giới hạn ở, số lượng đăng ký sử dụng; số lượng HKEY_LOCAL_MACHINE; số lượng HKEY_CLASSES_ROOT; có sử dụng các khoá đăng ký khởi tạo hay không; và

có sử dụng các khoá đăng ký cài đặt đang hoạt động hay không.

4. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 3, trong đó các đặc trưng về kết nối mạng được xác định trước có liên quan đến sự kiện kết nối mạng, bao gồm nhưng không giới hạn ở, tiến trình có kết nối mạng hay không; số lần kết nối mạng của tiến trình; địa chỉ IP, tên miền hoặc yêu cầu HTTP mà tiến trình kết nối tới; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không rõ nguồn gốc; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm có dấu hiệu nghi ngờ; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không hợp pháp; số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm đáng tin cậy; và số lượng các địa chỉ IP, tên miền, yêu cầu HTTP thuộc nhóm không an toàn.

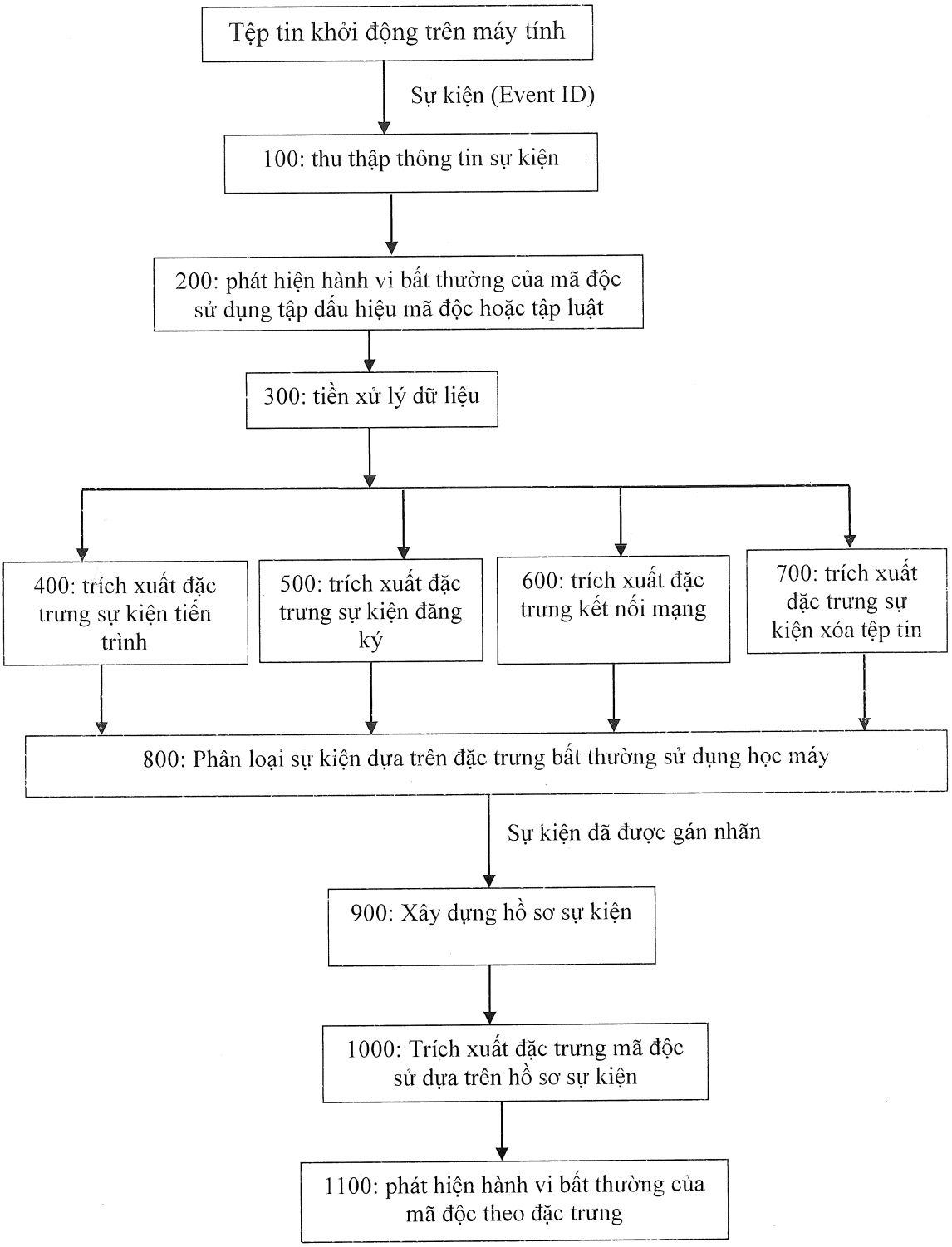
5. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 4, trong đó các đặc trưng về xoá tệp tin được xác định trước có liên quan đến sự kiện xóa tệp tin, bao gồm nhưng không giới hạn ở, số lượng các tệp tin tiến trình xóa; tổng dung lượng các tệp tin tiến trình xóa; có tiến trình xóa tệp tin ở ổ người dùng hay không; có tiến trình xóa tệp tin ở ổ $RECYCLE.BIN hay không; có tiến trình xóa tệp tin ở ổ chứa hệ điều hành hay không; có tiến trình xóa tệp tin ở ổ chứa dữ liệu hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình hay không; có tiến trình xóa tệp tin ở ổ chứa các tệp tin chương trình (x86) hay không; có tiến trình xóa tệp tin ở ổ chứa các Perflog hay không; có tiến trình xóa tệp tin ở ổ chứa thư mục OneDriveTemp hay không; có tiến trình xóa tệp tin ở ổ chứa $WinREAgent hay không; số lượng tệp tin xml tiến trình xóa; số lượng tệp tin ảnh tiến trình xóa; số lượng tệp tin HTML tiến trình xóa; số lượng tệp tin văn bản tiến trình xóa; và số lượng tệp tin thực thi tiến trình xóa.

6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó phương pháp này còn bao gồm phân loại các sự kiện dựa trên các đặc trưng được xác định trước có trong tệp thực thi, và tập hợp các đặc trưng được xác định trước có trong tệp thực thi này vào đồ thị biểu diễn mối quan hệ sự kiện theo thời đã nêu.

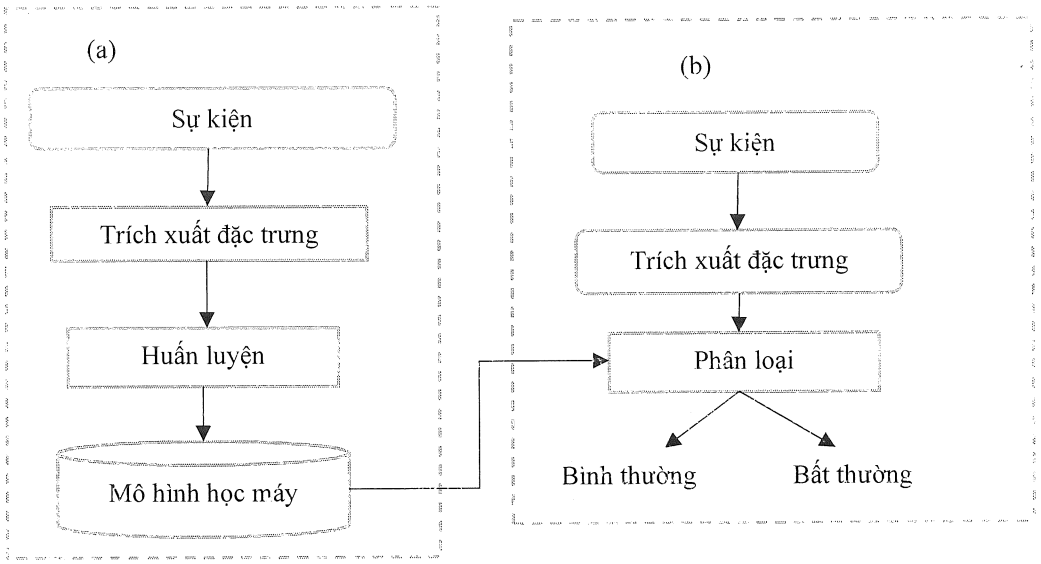
7. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 6, trong đó công cụ

phần mềm có chức năng thu thập dữ liệu trên máy tính người dùng là một trong số công cụ thu thập và lưu trữ sự kiện Sysmon, Wahzu, hoặc Endpoint Security được cài đặt theo gói phần mềm được biên dịch sẵn trên kho ứng dụng của Ubuntu.

8. Phương tiện lưu trữ có chứa chương trình máy tính bao gồm các lệnh mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 7.


Tệp tin khởi động trên máy tính
Sự kiện (Event ID)
100: thu thập thông tin sự kiện
200: phát hiện hành vi bất thường của mã độc sử dụng tập dấu hiệu mã độc hoặc tập luật
300: tiền xử lý dữ liệu
400: trích xuất đặc trưng sự kiện tiến trình
500: trích xuất đặc trưng sự kiện đăng ký
600: trích xuất đặc trưng kết nối mạng
700: trích xuất đặc trưng sự kiện xóa tệp tin
800: Phân loại sự kiện dựa trên đặc trưng bất thường sử dụng học máy
Sự kiện đã được gán nhãn
900: Xây dựng hồ sơ sự kiện
1000: Trích xuất đặc trưng mã độc sử dựa trên hồ sơ sự kiện
1100: phát hiện hành vi bất thường của mã độc theo đặc trưng

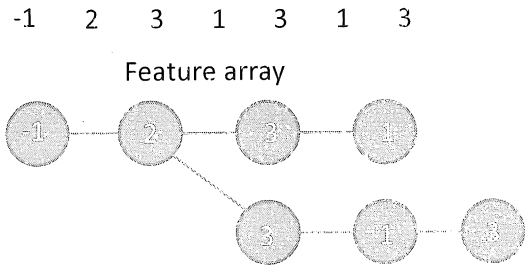
Hình 1



Hình 2

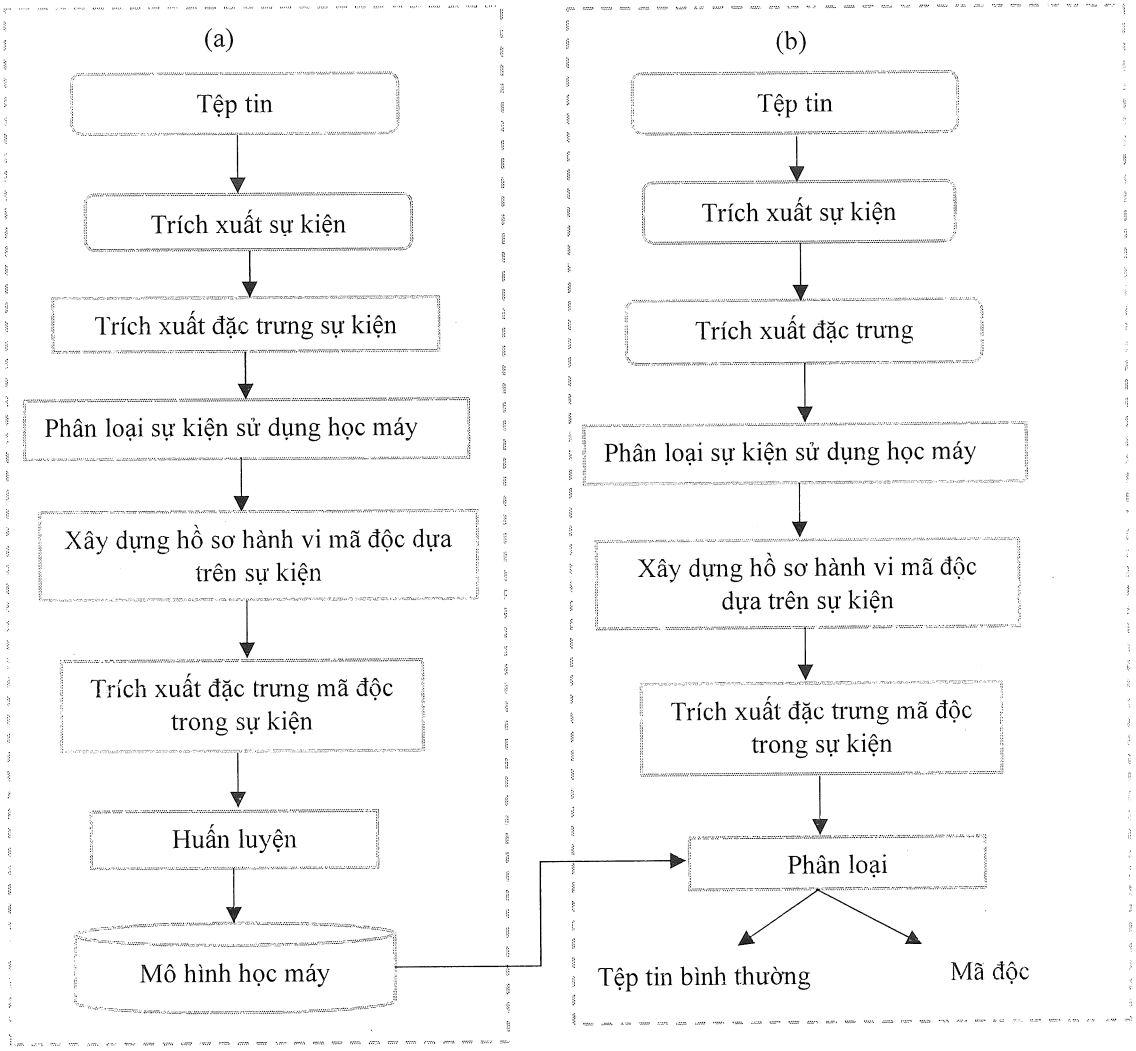
0	1	0	0	0	0	0
1	0	1	0	1	0	0
0	1	0	1	0	0	0
0	0	1	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	1	0	1
0	0	0	0	0	1	0

Adjacency matrix



Processes graph

Hình 3


(a)
Tệp tin
Trích xuất sự kiện
Trích xuất đặc trưng sự kiện
Phân loại sự kiện sử dụng học máy
Xây dựng hồ sơ hành vi mã độc dựa trên sự kiện
Trích xuất đặc trưng mã độc trong sự kiện
Huấn luyện
Mô hình học máy
(b)
Tệp tin
Trích xuất sự kiện
Trích xuất đặc trưng
Phân loại sự kiện sử dụng học máy
Xây dựng hồ sơ hành vi mã độc dựa trên sự kiện
Trích xuất đặc trưng mã độc trong sự kiện
Phân loại
Tệp tin bình thường
Mã độc

Hình 4